United Nations E/CN.3/2008/13

Economic and Social Council

Distr.: General
12 December 2007

Original: English

Statistical Commission
Thirty-ninth session
26-29 February 2008
Item 3 (l) of the provisional agenda*
Items for discussion and decision: common open standards for the exchange and sharing of data and metadata

Common open standards for the exchange and sharing of data and metadata

Note by the Secretary-General

In accordance with a request of the Statistical Commission at its thirty-eighth session,** the Secretary-General has the honour to transmit the present report on an initiative to foster common open standards for the exchange and sharing of data and metadata. The report is presented to the Commission for discussion.

The report reviews activities related to the Statistical Data and Metadata Exchange initiative and includes a proposal for its recognition as the international standard for data and metadata exchange.

The Commission may wish to express its views on progress achieved by the initiative and the way forward. Points for discussion by the Commission are contained in paragraph 20.

* E/CN.3/2008/1.

** See *Official Records of the Economic and Social Council, 2007, Supplement No. 4* (E/2007/24), chap. I.A.

07-63962 (E) 070108

Report on common open standards for the exchange and sharing of socio-economic data and metadata: the SDMX initiative*

I. Introduction

1. In 2001, the Bank for International Settlements, the European Central Bank, Eurostat, the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD) and the United Nations joined together to develop more efficient processes for the exchange and sharing of data and metadata within the current scope of their collective activities. The World Bank joined the initial group of sponsor organizations in 2003.

2. The goal of the Statistical Data and Metadata Exchange (SDMX) initiative is to foster standards and guidelines that allow national and international organizations to gain efficiencies and avoid duplication of work in the area of data and metadata exchange through the use of modern technology. The above sponsor organizations have been making progress over the past few years, especially through the increasing involvement of international and national statistical agencies. SDMX builds on existing and emerging technical exchange protocols and on the content-oriented efforts of statisticians who have worked on those long-standing issues in various domains and forums. More details can be found on the SDMX website (see http://www.sdmx.org).

3. The Statistical Commission received an initial report on the initiative during its thirty-third session in 2002.[1] A follow-up report was submitted to the Commission at its thirty-fourth session in 2003,[2] focusing on the launch of projects jointly undertaken by the sponsoring institutions. Subsequent reports in 2004[3] and 2005[4] highlighted further progress; and a special outreach event about SDMX was attended by about 50 delegates to the 2006 Commission session. The report in 2007[5] noted that SDMX technical standards and content-oriented guidelines had started to become a focal point for improving the exchange and sharing of statistical data and metadata in various national and international forums. It also noted that advances had reflected the orientation of sponsoring organizations to ensure open involvement of national and international statistical institutions in shaping future SDMX developments and implementations. The present report reviews SDMX activities and includes a proposal for recognition of SDMX as the international standard for data and metadata exchange. In particular, section II summarizes SDMX benefits, section III highlights SDMX implementations and costs, section IV takes up fostering of capacity-building and section V outlines conclusions and a recommendation to give recognition to SDMX, thereby fostering further

* The report has been jointly prepared by the Bank for International Settlements, the European Central Bank, the Statistical Office of the European Communities, the International Monetary Fund, the Organization for Economic Cooperation and Development, the United Nations Statistics Division and the World Bank.

[1] See *Official Records of the Economic and Social Council, 2002, Supplement No. 4* (E/2002/24), chap. VI. See also http://unstats.un.org/unsd/statcom/sc2002.htm.

[2] See http://unstats.un.org/unsd/statcom/sc2003.htm.

[3] See http://unstats.un.org/unsd/statcom/sc2004.htm.

[4] See http://unstats.un.org/unsd/statcom/sc2005.htm.

[5] See http://unstats.un.org/unsd/statcom/sc2007.htm.

collaboration and steps forward for the global community of statistical producers, disseminators and users.

II. Benefits of SDMX

4. SDMX consists of a number of products:

(a) Technical standards (information model, formats, architecture);[6]

(b) Content-oriented guidelines (cross-domain statistical concepts, subject-matter domains list, metadata common vocabulary);

(c) Implementation tools (e.g. creating data and metadata structures, transforming formats, data entry and registry services).

5. The benefits of SDMX have been spelled out in various reports for the Statistical Commission and for the Committee for the Coordination of Statistical Activities. To recapitulate briefly, SDMX products:

(a) Are applicable to all domains of statistics;

(b) Can improve efficiencies in the exchange and dissemination of data and metadata;

(c) Can be used by national and international agencies as key building blocks in internal statistical information technology systems used for the collection, compilation, storage and searching of statistical information;

(d) Are neutral in terms of underlying commercial technologies or their applicability in various statistical domains;

(e) Avoid duplication of efforts in developing and maintaining standards for the processing of statistical information;

(f) Can potentially reduce the cost of developing statistical software systems;

(g) Pool the expertise and resources working on data and metadata issues in sponsoring organizations and involve other experts at the national and international levels through an open consultation process;

(h) Are anchored in internationally recognized bodies such as the International Organization for Standardization for technical specifications and existing groups leading statistical methodologies for domain-specific content-oriented guidelines;

(i) Can be applied in various exchange modes, including those built on data "push" as well as "pull" models;

(j) Can potentially reduce duplication of reporting by facilitating data-sharing arrangements through data "hubs" or portals organized by several organizations in cooperation.

[6] Recognized by the International Organization for Standardization as Technical Specification 17369-SDMX.

6. Those benefits are real and are reflected in the many applications and projects that are in progress or are already making use of SDMX, at both national and international levels. The latter include:

(a) SDMX Open Data Interchange (Eurostat);

(b) Joint Comtrade Statistics System (Statistics Division and OECD);

(c) National Accounts World Wide Exchange (OECD);

(d) Joint External Debt Hub (Bank for International Settlements, IMF, OECD, World Bank);

(e) Eurosystem Joint Dissemination (European Central Bank and euro area national central banks);

(f) Statistical Data Warehouse (European Central Bank);

(g) Metadata Repositories (IMF);

(h) Education Statistics Hub (United Nations Educational, Scientific and Cultural Organization, OECD, Eurostat);

(i) CountrySTAT Hub for Agricultural Statistics (Food and Agriculture Organization of the United Nations);

(j) Labour Statistics Hub (International Labour Organization, OECD, Eurostat);

(k) Millennium Development Goals Hub (Statistics Division).

7. Recognition of SDMX achievements has also led to endorsement or adoption by the Committee for the Coordination of Statistical Activities, the European Union Statistical Programme Committee, the OECD Statistics Committee and the Statistics Committee of the European System of Central Banks. The Committee for the Coordination of Statistical Activities agreed that it would itself put forward SDMX to the Statistical Commission for consideration as a standard that it had adopted, stressing the importance of assisting countries to develop their capacity to adopt the standards. Support from the Statistical Programme Committee will be instrumental in achieving the expected benefits of using SDMX within the European Statistical System that have been identified. The OECD Statistical Committee agreed to support SDMX as the world standard for international exchange and sharing of statistical information. Following agreements by its Statistics Committee, the European System of Central Banks has been using SDMX in dissemination activities (e.g. Eurosystem Joint Dissemination) and, moreover, SDMX is already the unique standard used in statistical data-sharing between national central banks and the European Central Bank.

III. Implementations and costs of the Statistical Data and Metadata Exchange initiative

8. While the benefits of SDMX may be relatively easy to explain, questions have been raised as to the cost of implementing international standards and guidelines. National statistical agencies are understandably reluctant to embrace a new set of standards if they are costly to implement.

9. It should first be pointed out that the development, maintenance and dissemination of SDMX technical standards and guidelines is done by the sponsoring international organizations as a public good.[7] SDMX products are made available free of charge and without restrictions on use. The sponsoring organizations also foster the development and maintenance of free implementation tools, by public-sector as well as commercial organizations. Experts on statistical data and metadata exchange are invited to provide comments on, and suggestions for improving, the various SDMX products. Any major change in technical standards and content-oriented guidelines is carried out through an open consultation process announced on the SDMX website and carried forward by the sponsor organizations.

10. The full benefits of SDMX are achieved when they are complemented by domain-specific guidelines, i.e. data and metadata structure definitions and code lists for individual statistical domains such as national accounts, balance of payments, external debt or agricultural, labour and education statistics. SDMX fosters the development of such domain-specific implementations through existing international statistical groups that are responsible for international methodological standards in their respective statistical areas. It also promotes an open consultation process for the development and maintenance of those domain-specific SDMX implementations. The SDMX website will be providing an overview of that content-oriented work.

11. There are many ways in which international and national statistical agencies, as well as other compilers and users of statistical information in the private and public sector, can implement SDMX. In all cases they can be considered "business as usual" for technical experts and statisticians. Indeed, any statistical project will require the development of an information model and a structure for data and metadata; that is irrespective of whether the project involves the construction or adaptation of a database, data warehouse, data exchange platform or web application. To some extent it could be argued that all technical experts and statisticians have been using SDMX-like solutions in the past. SDMX provides essential building blocks for that work, based on a common language, shared experience and developing best practice, thereby potentially saving experts and statisticians considerable development and maintenance time.

12. The implementation of SDMX does not require a "big bang" approach and a need to change information technology systems or code structures. Implementation can be done gradually over time and at the pace dictated by institutions' own priorities. It is also possible to map existing solutions to SDMX standards or to design interfaces for existing systems, using a growing set of freely available tools to facilitate developments.

13. Early in 2007, the SDMX sponsoring institutions organized a global conference, hosted by the World Bank, to explain developments of SDMX and to illustrate its implementation by international and national agencies in the context of their own goals and priorities. The various case studies presented showed the gradual implementation of SDMX through a number of projects:

[7] A memorandum of understanding among the seven international sponsoring institutions was signed and published on the SDMX website (http://sdmx.org/?p=36) in 2007.

(a) Individual international organizations are introducing SDMX in their data exchange with national agencies for a variety of statistical domains;

(b) Groups of international organizations, through existing or new collaboration arrangements, are applying SDMX;

(c) Existing and new domain groups are working to develop SDMX-conformant data and metadata structures;

(d) National statistical agencies are turning to SDMX for their various projects.[8]

14. The cost of implementing SDMX is embedded in those different projects and it is therefore difficult to separate the actual net cost of introducing or using SDMX. Practical experiences have provided growing evidence to the effect that the relative costs for technology and subject-matter domains work are minor (as mentioned above, information models and data and metadata structure definitions need to be developed anyway). The most critical areas are seen as capacity-building for staff to better understand the SDMX framework and how to get full leverage out of applying it.

IV. Fostering capacity-building

15. So far, SDMX sponsoring organizations have promoted capacity-building through various initiatives:

(a) The posting of various technical manuals and user guidelines on the user-friendly SDMX website (http://www.sdmx.org);

(b) The provision of information on that website regarding other public-sector and commercial organizations making available free SDMX implementation tools (some tool providers offer discussion forums or practical guidelines);

(c) The posting of relevant information on SDMX implementations, including contact details, on the SDMX website (national and international statistical agencies are encouraged to keep the SDMX secretariat informed of implementation projects and to share information through the website);

(d) "Learning by doing" in the context of their own projects, individually or collaboratively,[9] which in many cases also involve their national constituencies in a practical way;

(e) The active marketing of SDMX through existing international groups and meetings at which SDMX sponsors are represented;

[8] Further details about presentations are available from the SDMX website (http://sdmx.org/?p=22#more-22), including country reports on developments in Australia, Mexico, the Philippines, the United States of America and Viet Nam. Other presentations provided an overview of SDMX developments, demonstrated practical implementations and provided introductory capacity-building in both technical and content-oriented aspects of SDMX. Nearly 200 technical and subject-matter experts from national and international organizations around the world participated.

[9] Some projects are so-called SDMX-recognized projects since they are identified as potentially yielding major improvements in SDMX products.

(f) The organization of occasional conferences and satellite workshops such as the global conference organized early in 2007 at the World Bank;

(g) The organization of training activities aimed at international and national statistical institutions dealing with the SDMX products and their domain-specific implementation.

16. The SDMX sponsors are also thinking of additional ways in which the implementation of SDMX could be promoted. The following will be considered:

(a) The organization of regional conferences and adjoining workshops along the lines of the global event organized early in 2007 in Washington, D.C.;

(b) The sponsoring of additional meetings hosted by national and international organizations to explain and illustrate SDMX in the context of their specific projects or domains in which they are actively involved;

(c) The development of new arrangements for SDMX capacity-building, for instance through the Partnership in Statistics for Development in the 21st Century and the Development Gateway Foundation (e.g. a virtual statistical system).

17. Those additional initiatives would be expected to attract greater participation if the Statistical Commission were to give recognition to SDMX achievements in fostering global standards and guidelines for statistical data and metadata exchange.

V. Conclusions

18. SDMX technical standards and content-oriented guidelines have demonstrated their potential for improving the exchange and sharing of statistical data and metadata using modern technology.

19. A number of implementations are in progress and endorsements of SDMX reflect a widening range of international and national statistical agencies.

20. The Commission may wish to discuss the achievements of SDMX and the way forward, with a view to giving recognition to SDMX as the standard for data and metadata exchange.

21. The Commission may wish to explicitly encourage efforts to:

(a) Further involve national and international agencies in their contribution of comments and requirements that can enhance the relevance of SDMX developments for statistical systems around the world;

(b) Continue outreach via the SDMX website, helping to broaden awareness-building and prevent duplication of efforts within the community of statistical producers, disseminators and users;

(c) Increase opportunities to benefit statistical systems and interchanges of data and metadata through freely available tools and common approaches;

(d) Foster capacity-building programmes, especially for developing countries, aimed at strengthening the know-how of technical experts and subject-matter domain experts, helping them to better understand SDMX and how its framework of technical standards and content-oriented guidelines can facilitate more effective and efficient statistical systems.